

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 15:09:02 ; Search time 22 Seconds

(without alignments)  
152.708 million cell updates/sec

Title: US-09-924-102-2

Sequence: 1 MLSTHFLFYFLFYFLSYSL.....RWGGGGRGRTADTGGMFLS 81

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	16.3	462	1	SR54_HAEIN
2	68	16.3	511	1	COPD_HUMAN
3	67	16.0	829	1	TOP1_XENLA
4	65	15.6	65	1	RPR_DROME
5	65	15.6	386	1	R887_DROME
6	64.5	15.4	365	1	R851_HUMAN
7	64	15.3	511	1	COPD_BOVIN
8	64	15.3	765	1	TOP1_HUMAN
9	64	15.3	767	1	TOP1_MOUSE
10	63.5	15.2	389	1	DNMT_HAICU
11	63.5	15.2	391	1	DNMT_HAICU
12	62	14.8	1898	1	TRHY_HUMAN
13	61.5	14.7	268	1	EP34_HCMVA
14	61.5	14.7	272	1	RS2_CAEEL
15	61.5	14.7	684	1	EP84_HCMVA
16	61.5	14.7	824	1	ROU_HUMAN
17	61	14.6	78	1	DMS1_PHYBI
18	61	14.6	404	1	LA_BOVIN
19	61	14.6	486	1	DNAB_HELPV
20	61	14.6	767	1	TOP1_CRIGR
21	61	14.6	876	1	SYA_SALTI
22	61	14.6	876	1	SYA_SALTY
23	59.5	14.2	165	1	LEI3_GOSHI
24	59.5	14.2	433	1	THC1_METTH
25	59.5	14.2	766	1	EYA_DROME
26	59	14.1	307	1	CC36_CAEEL
27	59	14.1	519	1	ELAY_DROVI
28	59	14.1	524	1	T2FA_XENLA
29	59	14.1	739	1	BCSA_PSEFL
30	59	14.1	1029	1	YK95_RHIME
31	58.5	14.0	140	1	GV11_HAICU
32	58.5	14.0	365	1	R851_MOUSE
33	58.5	14.0	1228	1	ECM_HUMAN

34	58.5	14.0	2254	1	CCAG_RAT	054898	rattus norv
35	58.5	14.0	2377	1	CCAG_HUMAN	043497	homo sapien
36	58	13.9	43	1	GRW1_LYCSC	001157	lyocopersico
37	58	13.9	79	1	YSZ1_CAEEL	017811	caenorhabd1
38	58	13.9	123	1	LSM4_CAEEL	019952	caenorhabd1
39	58	13.9	307	1	SP73_SCHPO	014311	schizosacch
40	58	13.9	327	1	FBRL_MOUSE	P35550	mus musculu
41	58	13.9	390	1	LA_DROME	P40796	drosophila
42	58	13.9	408	1	LA_HUMAN	P05455	homo sapien
43	58	13.9	409	1	R238_HUMAN	P54727	homo sapien
44	58	13.9	864	1	YCL8_HUMAN	Q9ULX2	homo sapien
45	58	13.9	943	1	YLW5_CAEEL	P34408	caenorhabd1

RESULT 1

SR54\_HAEIN STANDARD; PRT; 462 AA.

AC P44518;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Signal recognition particle protein (fifty-four homolog).

GN FPH OR H10106.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;

OC Haemophilus.

OX NCBI\_Taxid=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-R4 / KM20 / ATCC 51907;

RX MEDLINE=9530630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Keriavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shrivley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RL Science 269:496-512(1995).

CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE RIBOSOMES (BY SIMILARITY).

CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA MOLECULE AND PROTEIN FPH (BY SIMILARITY).

CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL SEQUENCE.

CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: U32696; AAC21784.1; -

CC HSSP: 007347; IFFH.

CC TIGR: H10106; -

CC InterPro: IPR003593; AAA\_ATPase.

CC InterPro: IPR000897; SRP54.

CC InterPro: IPR004125; SRP54\_SPB.

CC InterPro: IPR004780; SRP\_sub.

CC Pfam: PF00448; SRP54; 1.

CC Pfam: PF02881; SRP54\_N; 1.

```

DR Pfam: PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfam; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KM Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 295 G-DOMAIN.
FT DOMAIN 296 453 M-DOMAIN.
FT NP_BIND 107 114 GTP (BY SIMILARITY).
FT NP_BIND 190 194 GTP (BY SIMILARITY).
FT NP_BIND 248 251 GTP (BY SIMILARITY).
SQ SEQUENCE 462 AA; 50843 MW; 16AC32089A158F7A CRC64;

Query Match 16.3%; Score 68; DB 1; Length 462;
Best Local Similarity 27.9%; Pred. No. 3.7;
Matches 29; Conservative 14; Mismatches 31; Indels 30; Gaps 5;

QY 3 LSTH-----LFYLFYFLYSLSGDR-----RLCKRTQOQEQQLRQ 43
DB 360 LSEHVNQVDKRFVMEALINSMTLKERANDPDIINGRRRIALGSGTQVDVNLKQ 419
QY 44 SEVLEFRSETLRK-----TGKGRWGQGGGGGTADTGGF 79
DB 420 FDEMQRMMKMKRKGAKRMKRGQGLMGGLGSLGSG---LGGMF 460

RESULT 2
COPD_HUMAN
ID COPD_HUMAN STANDARD; PRT; 511 AA.
AC P48444;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coatomer delta subunit (Delta-coat protein) (Delta-COP) (Archain).
GN COPD OR ARCN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95301274; PubMed=7782067;
RA Radice P., Pensotti V., Jones C., Perry H., Pierotti M.A.,
RA Tunaci A.;
RT "The human archain gene, ARCN1, has highly conserved homologs in rice
RT and Drosophila."
RL Genomics 26:101-106(1995).
CC -1- FUNCTION: THE COATOMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
CC TO DIVISIVE MOTORS AND REVERSIBLY ASSOCIATES WITH GOLGI NON-
CC CLATHRIN-COATED VESICLES, WHICH FURTHER MEDIATE BIOSYNTHETIC
CC PROTEIN TRANSPORT FROM THE ER, VIA THE GOLGI UP TO THE TRANS GOLGI
CC NETWORK. COATOMER COMPLEX IS REQUIRED FOR BUDDING FROM GOLGI
CC MEMBRANES, AND IS ESSENTIAL FOR THE RETROGRADE GOLGI-TO-ER
CC TRANSPORT OF DILYSINE-TAGGED PROTEINS. IN MAMMALS, THE COATOMER
CC CAN ONLY BE RECRUITED BY MEMBRANES ASSOCIATED TO ADP-RIBOSYLATION
CC FACTORS (ARFES), WHICH ARE SMALL GTP-BINDING PROTEINS; THE COMPLEX
CC ALSO INFLUENCES THE GOLGI STRUCTURAL INTEGRITY, AS WELL AS THE
CC PROCESSING, ACTIVITY, AND ENDOCYTIC RECYCLING OF LDL RECEPTORS (BY
CC SIMILARITY).
CC -1- SUBUNIT: OLIGOMERIC COMPLEX THAT CONSISTS OF AT LEAST THE ALPHA,
CC BETA, BETA', GAMMA, DELTA, EPSILON AND ZETA SUBUNITS.
CC -1- SUBCELLULAR LOCATION: THE COATOMER IS CYTOPLASMIC OR POLYMERIZED
CC ON THE CYTOPLASMIC SIDE OF THE GOLGI, AS WELL AS ON THE
CC VESICLES/BUDS ORIGINATING FROM IT (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES MEDIUM SUBUNITS
CC FAMILY. COPD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81197; CAA57071.1; -.
DR EMBL; X81198; CAA57072.1; ALT_INIT.
DR Genew; HGNC:649; ARCN1.
DR MIM; 600820; -.
DR InterPro; IPR001392; Clathrn.med.
DR Pfam; PF00928; Adap_comp_sub; 1.
KM Transport; Protein transport; Golgi stack; Membrane; Polymorphism.
FT VARIANT 186 186 F->L (IN DBSNP:682327).
FT VARIANT 309 309 F->N (IN DBSNP:1063124).
SQ SEQUENCE 511 AA; 57210 MW; 4ED1F72D12A7F75 CRC64;

Query Match 16.3%; Score 68; DB 1; Length 511;
Best Local Similarity 21.1%; Pred. No. 4.1;
Matches 20; Conservative 24; Mismatches 29; Indels 22; Gaps 2;

QY 3 LSTHLEFYFLYSLSGDRRL-----CLRTKQOQEQQLR 42
DB 102 ISEHCFDLFADEYALGIRENVNLAQRTPEMDSHEKYFRAVRETQERAKEMR 161
QY 43 OSEVLEFRSETLRKTKGRWGQGGGGGTADTGG 77
DB 162 KAKEL--QOARDAERQGRKAPGFGGSSAVSGG 194

RESULT 3
TOP1_XENLA
ID TOP1_XENLA STANDARD; PRT; 829 AA.
AC P41512;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2).
GN TOP1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96433160; PubMed=8836188;
RA Pandit S.D., Richard R.E., Sternglanz R., Bogenhagen D.F.;
RT "Cloning and characterization of the gene for the somatic form of DNA
RT topoisomerase I from Xenopus laevis."
RL Nucleic Acids Res. 24:3593-3600(1996).
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SWEPT DNA STRAND.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: L07777; AAB36608.1; -
DR HSSP: P11387; I435.
DR InterPro: IPR001631; Topoisomerase_I.
DR Pfam: PF01028; Topoisomerase_I.1.
DR Pfam: PF02819; Topoisomerase_I.1.
DR PRINTS: PR00416; EUPISMRASE1.
DR SMART: SM00435; TOPEUC; 1.
DR PROSITE: PS00176; TOPOISOMERASE_I_EUK; 1.
DR Isomerase: Topoisomerase; DNA-binding.
FT ACT SITE 779 779 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 829 AA; 98230 MW; 8D1FE4252A910219 CRC64;

Query Match 16.0%; Score 67; DB 1; Length 829;
Best Local Similarity 30.0%; Pred. No. 8.8;
Matches 18; Conservative 9; Mismatches 33; Indels 0; Gaps 0;

QY 18 YSLGDRALCLRTKQOKEQOILROSEVYLRSELTFRRTGKGRNRGSGGGRGATDNGC 77
DB 500 YETARLKMCEYKINRTYKEDWKSKEKVRORAVALYFDIKIALAGNEKEGEGTADTVG 559

RESULT 4
RPR_DROME STANDARD; PRT; 65 AA.
AC Q24475; Q9VVP7;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Reaper protein.
GN RPR OR CG4319.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=EYE imaginal disk;
RX MEDLINE=9423205; PubMed=8171319;
RT White K., Grether M.E., Abrams J.M., Young L., Farrell K., Steller H.;
RL "Genetic control of programmed cell death in Drosophila.";
Science 264:677-683(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agdayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idbegwan C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Klamko B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshnefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 267:2185-2195(2000).
CC -1- FUNCTION: PLAYS A CENTRAL AND GLOBAL REGULATORY FUNCTION FOR THE
CC INITIATION OF APOPTOSIS. ECTOPIC EXPRESSION IN THE DEVELOPING EYE
CC RESULTS IN A SMALL EYE OWING TO EXCESS CELL DEATH.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION CORRESPONDS TO THE PATTERN OF
CC PROGRAMMED CELL DEATH IN THE EMBRYO.
CC -1- SIMILARITY: LIMITED AT THE N-TERMINAL, TO HID AND GRIM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L31631; AAL18983.1; -
DR EMBL: AE003520; AAF49264.1; -
DR DR Flybase: FBgn0011706; rpr.
KM Apoptosis.
SQ SEQUENCE 65 AA; 7682 MW; 57F231379AFEAE3C CRC64;

Query Match 15.6%; Score 65; DB 1; Length 65;
Best Local Similarity 37.7%; Pred. No. 0.96;
Matches 23; Conservative 11; Mismatches 11; Indels 16; Gaps 4;

QY 16 LYSIGDRARLCLRTKQOKEQOILROSEVYLR-----SETLR-----KTKKGR 62
DB 3 VAFIYPDQATL-LREA--EQKEQILRLRESQWRRLAVLLETILQRYTSCHPKGRKSGK 59

QY 63 W 63
DB 60 Y 60

RESULT 5
R87_DROME STANDARD; PRT; 386 AA.
AC Q24810;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein 87F (HRP36.1 protein) (P11
DE protein).
GN HRP87F OR HRP36.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R and Canton-S; TISSUE=Ovary;
RX MEDLINE=91187645; PubMed=18492577;
RA Haynes S.R., Johnson D., Raychaudhuri G., Beyer A.L.;
RT "The Drosophila HRP87F gene encodes a new member of the A and B hnRNP
RL protein group.";
RL Nucleic Acids Res. 19:25-31(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=92112968; PubMed=1730754;
RA Matunis E.L., Matunis M.J., Dreyfuss G.;

```

RT "Characterization of the major hnRNP proteins from Drosophila  
RT melanogaster.";  
RL J. Cell Biol. 116:257-269(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Canton-S; TISSUE=Embryo;  
RA MEDLINE=92020124; PubMed=1717937;  
RX Howeman B.T., Dessen E., Mechler H., Mack E.;  
RT "Drosophila snRNP associated protein p11 which specifically binds to  
RT heat shock puff 93D reveals strong homology with hnRNP core protein  
RT A1.";  
RL Nucleic Acids Res. 19:4909-4914(1991).  
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF RIBONUCLEOSOMES. COULD BE  
CC NEEDED TO ORGANIZE A CONCENTRATION GRADIENT OF A DORSALIZING  
CC MORPHOGEN (DM) ORIGINATING IN THE GERMAL VESICLE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND/OR CYTOPLASMIC.  
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RNM).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X54803; CAA38574.1; -;  
DR EMBL; X62636; CAA44502.1; -;  
DR EMBL; X59691; CAA42212.1; -;  
DR HSSP; P09651; IUP1.  
DR FLYBase; FBgn0004237; Hrb87F.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; Rrm; 2.  
DR SMART; SM00360; RRM; 2.  
DR PROSITE; PSS0102; RRM; 2.  
DR PROSITE; PSS0030; RRM\_RNP\_1; 2.  
KW RNA-binding; Nuclear protein; Ribonucleoprotein; Repeat;  
KW Alternative splicing  
FT DOMAIN 24 101 RNA-BINDING (RRM) 1.  
FT DOMAIN 115 192 RNA-BINDING (RRM) 2.  
FT VAAPSPIC 315 374 MISSING (IN ISOFORM HRP36.1).  
FT CONFLICT 271 271 S -> T (IN REF. 3).  
SQ SEQUENCE 386 AA; 39557 MW; 2036C04D1E3AF7 CRC64;  
Query Match 15.6%; Score 65; DB 1; Length 386;  
Best Local Similarity 50.0%; Pred. No. 6.5;  
Matches 13; Conservative 3; Mismatches 8; Indels 2; Gaps 1;  
OY 54 RRTGKGRMRWGOGGGRGTADTGGMF 79  
DB 231 RQNG--GGNWGAGAGGGGFGNSGNNF 254  
RESULT 6  
RRSL\_HUMAN STANDARD; PRT; 365 AA.  
ID RRSL\_HUMAN  
AC Q15050; O9BUX8; 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribosome biogenesis regulatory protein homolog.  
GN RRR OR KIAA0112.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=95308325; PubMed=7788527;  
RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,  
RA Tabata S., Ishikawa K.-I., Kawarayashi Y., Kotani H., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. III.

RT The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by  
RT analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 2:37-43(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung, and Uterus;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SUBCELLULAR LOCATION:  
RA Scherl A., Couté Y., Deon C., Sanchez J.-C., Diaz J.-J.,  
RA Hochstrasser D.;  
RT "Proteomic analysis of the nucleolus.";  
RL (in) Palagi P.M., Sanchez J.-C., Stoecklin R. (eds.);  
RL Proceedings of the Swiss Proteomic Society; 2001 Congress Functional  
RL Proteomics, pp.124-124, Fontis Media, Lausanne (2001).  
CC -1- FUNCTION: Involved in ribosome biogenesis (by similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -1- SIMILARITY: BELONGS TO THE RRS1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D25218; BAA04948.1; ALT\_INT.  
DR EMBL; BC001811; AAH01811.1; -;  
DR EMBL; BC013043; AAH13043.1; -;  
DR SWISS-2DPAGE; Q15050; HUMAN.  
KW Ribosome biogenesis; Nuclear protein.  
FT DOMAIN 302 365 ARG/GUY/LYS-RICH.  
SQ SEQUENCE 365 AA; 41193 MW; 061C3A6174C2ETC2 CRC64;  
Query Match 15.4%; Score 64.5; DB 1; Length 365;  
Best Local Similarity 37.5%; Pred. No. 6.9;  
Matches 18; Conservative 6; Mismatches 17; Indels 7; Gaps 2;  
OY 29 RRTGKGRMRWGOGGGRGTADTGGMF 76  
DB 286 RAINKQMRD---QEFAAR-----RMSQGRKRGGRGPGKRRKG 326  
RESULT 7  
COPD\_BOVIN STANDARD; PRT; 511 AA.  
ID COPD\_BOVIN  
AC P53619; 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coatomer delta subunit (Delta-coat protein) (Delta-COP).  
GN COPD.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OX Bovidae; Bovinae; Bos.  
OX NCBI\_Taxid=9913;  
RN [1]  
RP SEQUENCE OF 2-511 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97011152; PubMed=8858162;  
RA Paulstich D., Auerbach S., Orci L., Ravazzola M., Wegehingel S.,  
RA Lottspeich F., Stenbach G., Harter C., Wieland F.T., Tschöner H.;  
RT "Architecture of coatomer: molecular characterization of delta-COP  
RT and protein interactions within the complex.";  
RL J. Cell Biol. 135:53-61(1996).  
CC -1- FUNCTION: THE COATOMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS  
CC TO DILYSINE MOTIFS AND REVERSIBLY ASSOCIATES WITH GOLGI NON-  
CC CLATHRIN-COATED VESICLES, WHICH FURTHER MEDIATE BIOSYNTHETIC  
CC PROTEIN TRANSPORT FROM THE ER, VIA THE GOLGI UP TO THE TRANS GOLGI  
CC NETWORK. COATOMER COMPLEX IS REQUIRED FOR BUDDING FROM GOLGI



CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: J03250: AAA61207.1; -  
 DR EMBL: M60706: AAA61206.1; JOINED.  
 DR EMBL: M60688: AAA61206.1; JOINED.  
 DR EMBL: M60689: AAA61206.1; JOINED.  
 DR EMBL: M60690: AAA61206.1; JOINED.  
 DR EMBL: M60691: AAA61206.1; JOINED.  
 DR EMBL: M60692: AAA61206.1; JOINED.  
 DR EMBL: M60693: AAA61206.1; JOINED.  
 DR EMBL: M60694: AAA61206.1; JOINED.  
 DR EMBL: M60695: AAA61206.1; JOINED.  
 DR EMBL: M60696: AAA61206.1; JOINED.  
 DR EMBL: M60697: AAA61206.1; JOINED.  
 DR EMBL: M60698: AAA61206.1; JOINED.  
 DR EMBL: M60699: AAA61206.1; JOINED.  
 DR EMBL: M60700: AAA61206.1; JOINED.  
 DR EMBL: M60701: AAA61206.1; JOINED.  
 DR EMBL: M60702: AAA61206.1; JOINED.  
 DR EMBL: M60703: AAA61206.1; JOINED.  
 DR EMBL: M60704: AAA61206.1; JOINED.  
 DR EMBL: M60705: AAA61206.1; JOINED.  
 DR EMBL: U07804: AAB60379.1; -  
 DR EMBL: U07806: AAB60380.1; -  
 DR EMBL: A1035652: CAB43980.1; -  
 DR EMBL: X16479: CAA34500.1; ALU\_INT.  
 DR EMBL: M27913: AAA61208.1; -  
 DR PIR: A30887: ISHUT1.  
 DR PDB: 1A31: 19-AUG-98.  
 DR PDB: 1A35: 26-AUG-98.  
 DR PDB: 1A36: 12-AUG-98.  
 DR Aarhus/Ghent-2DPAGE: 610; NEPHE.  
 DR Genew: HGNC:11986; TOP1.  
 DR MIM: 126420; -  
 DR InterPro: IPR001631; Topisomerase-I.  
 DR Pfam: PF01028; Topisomerase-I; 1.  
 DR Pfam: PF02919; Topisomerase-I; 1.  
 DR PRINTS: PR00416; EUTPISMRASE1.  
 DR SMART: SM00435; TOPEUC; 1.  
 DR PROSITE: PS00176; TOPOISOMERASE\_I\_EUK; 1.  
 DR KIM: Topoisomerase; DNA-binding; Polymorphism; 3d-structure.  
 DR FT: 191 197  
 DR DOMAIN ACT\_SITE 723 723  
 DR FT: 370 370  
 DR FT: 370 370  
 DR FT: 533 533  
 DR FT: 722 722  
 DR FT: 729 729  
 DR FT: 145 145  
 DR FT: 145 145  
 DR FT: 765 AA: 90725 MM; 6FBD540BCF7BE28 CRC64;  
 DR SQ: 765 AA: 90725 MM; 6FBD540BCF7BE28 CRC64;  
 DR Query Match 15.3%; Score 64; DB 1; Length 765;  
 DR Best Local Similarity 30.0%; Pred. No. 17;  
 DR Matches 18; Conservative 9; Mismatches 33; Indels 0; Gaps 0;  
 DR 18 YSLGDRARLCTKRTKQKQEQQLRQSEVLFNSETLRKTKGKRRWGCGGCGGTADTGG 77  
 DR 444 YETARLLKCKVDKIRNRYREDKSKEMKRAVALYFDKLLALRGNENKESGETADTVG 503

DR 01-OCT-1993 (Rel. 27, last sequence update)  
 DR 16-OCT-2001 (Rel. 40, last annotation update)  
 DE DNA topoisomerase I (EC 5.99.1.2).  
 GN TOP1 OR TOP-1  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=93216125; PubMed=8096488.  
 RA Kawai O., Yasui Y., Sakai Y., Watanabe T., Ishii K.,  
 RA Yanagihara S., Andoh T.,  
 RT Cloning of the mouse cDNA encoding DNA topoisomerase I and  
 RL chromosomal location of the gene.  
 RL Gene 125:211-216(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hui C.F., Lo C.K., Hwang J.,  
 RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
 CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
 CC DNA, followed by passage and rejoining.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND IT CAN RELAX BOTH  
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES  
 CC RELAX ONLY NEGATIVE SUPERCOILS.  
 CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN  
 CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.  
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: D10061: BAA00950.1; -  
 DR EMBL: L20632: AAA40466.1; -  
 DR PIR: J01044; J01044.  
 DR HSSP: P11387; 1A35.  
 DR MGD: MGI:98788; TOP1.  
 DR InterPro: IPR001631; Topisomerase-I.  
 DR Pfam: PF01028; Topisomerase-I; 1.  
 DR Pfam: PF02919; Topisomerase-I; 1.  
 DR PRINTS: PR00416; EUTPISMRASE1.  
 DR SMART: SM00435; TOPEUC; 1.  
 DR PROSITE: PS00176; TOPOISOMERASE\_I\_EUK; 1.  
 DR KIM: Topoisomerase; DNA-binding.  
 DR FT: 725 725  
 DR ACT\_SITE 91 91  
 DR FT: 121 121  
 DR FT: 129 129  
 DR FT: 161 161  
 DR FT: 167 167  
 DR FT: 277 277  
 DR FT: 292 292  
 DR FT: 522 522  
 DR FT: 533 533  
 DR FT: 762 762  
 DR SQ: 767 AA: 90789 MM; 398327062B179F2A CRC64;  
 DR Query Match 15.3%; Score 64; DB 1; Length 767;  
 DR Best Local Similarity 30.0%; Pred. No. 17;  
 DR Matches 18; Conservative 9; Mismatches 33; Indels 0; Gaps 0;  
 DR 18 YSLGDRARLCTKRTKQKQEQQLRQSEVLFNSETLRKTKGKRRWGCGGCGGTADTGG 77

Query Match 15.28; Score 63.5; DB 1; Length 389;

```

CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005002; AAG19025.1; -
DR HSSP; P08622; 1BQ2.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXXCXXGXG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF00684; DnaJ_CXXCXXGXG; 1.
DR Pfam; PF01556; DnaJ_C; 1.
DR PRINTS; PRO0625; DNAJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00635; DnaJ_1; FALSE_NEG.
DR PROSITE; PS00766; DnaJ_2; 1.
DR PROSITE; PS00637; DnaJ_CXXCXXGXG; FALSE_NEG.
DR Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;
KW Complete proteome.
FT DOMAIN 4 68 GY-DOMAIN.
FT REPEAT 79 131 GLY-RICH.
FT REPEAT 165 172 CXXCXXGXG MOTIF.
FT REPEAT 182 189 CXXCXXGXG MOTIF.
FT REPEAT 208 215 CXXCXXGXG MOTIF.
FT REPEAT 222 229 CXXCXXGXG MOTIF.
FT METAL 165 165 ZINC 1 (BY SIMILARITY).

```

FT METAL 168 168 ZINC 1 (BY SIMILARITY).  
 FT METAL 182 182 ZINC 2 (BY SIMILARITY).  
 FT METAL 185 185 ZINC 2 (BY SIMILARITY).  
 FT METAL 208 208 ZINC 2 (BY SIMILARITY).  
 FT METAL 211 211 ZINC 2 (BY SIMILARITY).  
 FT METAL 222 222 ZINC 1 (BY SIMILARITY).  
 FT METAL 225 225 ZINC 1 (BY SIMILARITY).  
 SQ SEQUENCE 391 AA; 41713 MW; 3FFPD7E01B5EA25F CRC64;  
 Query Match 15.2%; Score 63.5; DB 1; Length 391;  
 Best Local Similarity 32.7%; Pred. No. 9.6;  
 Matches 18; Conservative 5; Mismatches 27; Indels 5; Gaps 1;  
 QY 30 KTKOQKREQQIL-----ROSEVLFRESELTGKGRKRRMGCGGREGTADTGGMF 79  
 DB 44 KFKKIKRQKVDLDEFTROOYDOLGHERFEAEKRGATGCGGAGMGAGGPF 98  
 RESULT 12  
 TRHY\_HUMAN STANDARD; PRT; 1898 AA.  
 ID 007283;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trichohyalin.  
 GN THH OR TRHY OR THL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93280194; PubMed=7685034;  
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,  
 RA Steinhert P.M.;  
 RT "The structure of human trichohyalin. Potential multiple roles as a  
 RT functional EF-hand-like calcium-binding protein, a cornified cell  
 RT envelope precursor, and an intermediate filament-associated (cross-  
 RT linking) protein.";  
 RT J. Biol. Chem. 268:12164-12176(1993).  
 RN [2]  
 RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=93315897; PubMed=7686953;  
 RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinhert P.M.;  
 RA "Trichohyalin: a structural protein of hair, tongue, nail, and  
 RA epidermis.";  
 RT J. Invest. Dermatol. 101:658-715(1993).  
 RL J. Invest. Dermatol. 101:658-715(1993).  
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
 CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
 CC DIFFERENTIATION.  
 CC -1- SUBUNIT: MONOMER (PROBABLE).  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING-TISSUES SUCH AS  
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN  
 CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF  
 CC THE EPIDERMIS.  
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND  
 CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST  
 CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS  
 CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED  
 CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS  
 CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.  
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN  
 CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG  
 CC DIFFERENT SPECIES.  
 CC -1- PTM: SUBSTATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE

CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L09190; AA65582.1; -  
 CC PIR: A45973; A45973.  
 CC HSP: P02633; AICB.  
 CC Genew: HGNC:11791; THH.  
 CC MIM: 190370; -  
 CC InterPro: IPR001751; CAPB\_S100.  
 CC InterPro: IPR002048; EF-hand.  
 CC InterPro: IPR002017; Spectrin.  
 CC Pfam: PF00036; efhand; 1.  
 CC Pfam: PF01023; S100; 1.  
 CC Pfam: PF003407; CAPB\_S100; 1.  
 CC PROSITE: PS00018; EF\_HAND; 1.  
 CC PROSITE: PS00303; S100\_CAPB; 1.  
 CC Kewatization; Repeat; Calcium-binding.  
 CC DOMAIN 1 91  
 CC CA\_BIND 22 33  
 CC CA\_BIND 62 73  
 CC DOMAIN 314 390  
 CC FT REPEAT 314 326  
 CC FT REPEAT 327 339  
 CC FT REPEAT 340 351  
 CC FT REPEAT 352 364  
 CC FT REPEAT 365 377  
 CC FT REPEAT 378 390  
 CC FT DOMAIN 391 444  
 CC FT REPEAT 391 396  
 CC FT REPEAT 397 402  
 CC FT REPEAT 403 408  
 CC FT REPEAT 409 414  
 CC FT REPEAT 415 420  
 CC FT REPEAT 421 426  
 CC FT REPEAT 427 432  
 CC FT REPEAT 433 438  
 CC FT REPEAT 439 444  
 CC FT DOMAIN 444 702  
 CC FT REPEAT 923 952  
 CC FT REPEAT 923 952  
 CC FT REPEAT 923 952  
 CC FT REPEAT 953 982  
 CC FT REPEAT 983 1012  
 CC FT REPEAT 1013 1042  
 CC FT REPEAT 1043 1072  
 CC FT REPEAT 1073 1102  
 CC FT REPEAT 1103 1132  
 CC FT REPEAT 1133 1162  
 CC FT REPEAT 1162 1192  
 CC FT DOMAIN 1250 1849  
 CC FT CONFLICT 1752 1752  
 CC FT CONFLICT 1794 1801  
 CC FT CONFLICT 1857 1857  
 CC FT CONFLICT 1880 1880  
 CC SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;  
 Query Match 14.8%; Score 62; DB 1; Length 1898;  
 Best Local Similarity 31.7%; Pred. No. 76;  
 Matches 13; Conservative 13; Mismatches 15; Indels 0; Gaps 0;  
 QY 22 DRARLCIRTKOQKREQQILROSEVLFRESELTGKGRKRR 62  
 DB 333 ERREQQLRRQGERRRRQQLRRQGERRRRQQLRRQGERRRR 373

```

RESULT 13
EP34_HCMVA STANDARD; PRT; 268 AA.
ID EP34_HCMVA
AC P16768;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Early phosphoprotein P34.
GN Utl12.
OS Human cytomegalovirus (strain AD169).
CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohnl R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Predde E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.,
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RT Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RL [2]
RN SEQUENCE FROM N.A.
RP Wang S.K., Duh C.Y.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-6 AND
CC HSV-7 U79 AND HCMV Utl12 (P34).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
DR EMBL: X17403; CA35315.1; ALT_TERM.
DR EMBL: U57433; AAB53251.1; -.
DR PIR: S09860; S09860.
DR InterPro: IPR004138; U79_P34.
DR Pfam: PF03064; U79_P34; 1.
KW Early protein; Phosphorylation.
FT DOMAIN 148 153 POLY-GLY.
FT DOMAIN 197 200 POLY-SER.
FT DOMAIN 202 220 POLY-CYS.
SQ SEQUENCE 268 AA; 28142 MW; BD5906FAEAB28FC9 CRC64;

Query Match 14.7%; Score 61.5; DB 1; Length 268;
Best Local Similarity 40.5%; Pred. No. 11;
Matches 17; Conservative 2; Mismatches 20; Indels 3; Gaps 1;

OY 36 KEOQILROSEVLFSEFLRTGKGRWGCGGGRGTADTGG 77
Db 179 EEOQRRROEO---RHEERRKKSSSSAGGGGCGAGGGGGG 217

RESULT 14
RS2_CAEL STANDARD; PRT; 272 AA.
ID RS2_CAEL
AC P51403;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S2.
GN RPS-2 OR C49H3.11.
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
CC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-Bristol N2;
RA Wu X.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
DR EMBL: U42436; AAF99899.1; -.
DR HSSP: P02357; LKP.
DR WormPep; C49H3.11; CE04237.
DR InterPro: IPR000831; Ribosomal_S5.
DR Pfam: PF00333; Ribosomal_S5_C.1.
DR Pfam: PF03719; Ribosomal_S5_C.1.
DR TIGRFAMs: TIGR01020; rpse_arch.1.
DR PROSITE: PS00585; RIBOSOML_S5; 1.
KW Ribosomal protein; Repeat.
SQ SEQUENCE 272 AA; 28960 MW; 60049F657DB8CB34 CRC64;

Query Match 14.7%; Score 61.5; DB 1; Length 272;
Best Local Similarity 56.0%; Pred. No. 11;
Matches 14; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 54 RKTGKGRW-GGCGRGRTADTGG 77
Db 27 RPAGRGGRGGRGGRGGRGGRG 51

RESULT 15
EP34_HCMVA STANDARD; PRT; 684 AA.
ID EP34_HCMVA
AC P17151; P87888; O12567;
DT 01-AUG-1990 (Rel. 15, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Early phosphoprotein P84.
GN Utl12/Utl13.
OS Human cytomegalovirus (strain AD169).
CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohnl R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Predde E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.,
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RT Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RN [2]
RP REVISIONS, AND IDENTIFICATION.
RA Wang S.K., Duh C.Y.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
DR EMBL: X17403; CA35315.1; -.
DR EMBL: U57432; AAB53251.1; -.
DR PIR: S09880; S09880.
DR InterPro: IPR004138; U79_P34.

```

DR Pfam; PF03064; U79\_P34; 1.  
KW Early protein; Phosphorylation.  
FT DOMAIN 148; 153 POLY-GLY.  
FT DOMAIN 197 200 POLY-SER.  
FT DOMAIN 202 220 POLY-GLY.  
FT DOMAIN 271 276 POLY-GLY.  
FT DOMAIN 285 296 POLY-GLY.  
FT DOMAIN 314 318 POLY-SER.  
FT DOMAIN 321 324 POLY-SER.  
FT DOMAIN 327 334 POLY-HIS.  
FT DOMAIN 390 397 POLY-ALA.  
FT DOMAIN 563 568 POLY-SER.  
FT DOMAIN 576 576 POLY-SER.  
FT DOMAIN 578 585 POLY-PRO.  
FT DOMAIN 599 606 POLY-GLY.  
FT DOMAIN 615 621 POLY-SER.  
SQ SEQUENCE 684 AA; 70272 MW; F3325CE2E29BF449 CRC64;  
Query Match 14.78; Score 61.5; DB 1; Length 684;  
Best Local Similarity 40.58; Pred. No. 29;  
Matches 17; Conservative 2; Mismatches 20; Indels 3; Gaps 1;  
OY 36 KEOQILROSEVLFRRSEFLRKTGKKRRRWGGGGRGGTADTGG 77  
Db 179 EEOQRRRQEQ--RHERRRKKSSSSAGGGGGGAGGGGGGG 217

Search completed: May 29, 2003, 15:18:41  
Job time : 24 secs